AMENDMENTS TO THE SPECIFICATION

Please amend the paragraphs [006], [0011], [0014], [0018], [0019], [0022], [0023], [0027], [0028], [0029], [0030], [0031], [0032], [0035], [0038], [0043], [0045], [0050], [0055] and Tables 1 to 4 as follows:

[006] In one embodiment, a composition is provided that comprises a protein in crystalline form wherein the protein has 65%, 70%, 80%, 90%, 95% or greater identity with residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3.

[0011] In one embodiment, a method is provided for forming crystals of a protein comprising: forming a crystallization volume comprising: a protein that has at least 65%, 70%, 80%, 90%, 95% identity with residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3 in a concentration between 1 mg/ml and 50 mg/ml; 5-50% w/v of precipitant wherein the precipitant comprises one or more members of the group consisting of PEG MME having a molecular weight range between 300-10000, and PEG having a molecular weight range between 100-10000; optionally 0.05 to 0.8M additives wherein the additives comprises sarcosine or 0.5 to 25% additives wherein the additives comprises xylitrol; and wherein the crystallization volume has a pH between pH 5 and pH 9; and storing the crystallization volume under conditions suitable for crystal formation. The method optionally further comprises using 0.05-0.2M buffers selected from the group consisting of Tris-HCl, bicine and combinations thereof. The method also optionally further includes performing the crystallization at a temperature between 1°C - 25°C.

[0014] In one embodiment, machine readable data storage medium is provided having data storage material encoded with machine readable data, the machine readable data comprising: structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on alpha-carbon atoms positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3.

[0018] In one embodiment, a method is provided for displaying a three dimensional representation of a structure of a protein comprising: taking machine readable data comprising structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-

carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3; computing a three dimensional representation of a structure based on the structure coordinates; and displaying the three dimensional representation.

[0019] In another embodiment, a method is provided for displaying a three dimensional representation of a structure of a protein comprising: displaying a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates of Figure 3 that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3.

[0022] In one embodiment, a computational method is provided comprising: taking machine readable data comprising structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3; computing phases based on the structural coordinates; computing an electron density map based on the computed phases; and determining a three-dimensional crystal structure based on the computed electron density map.

[0023] In another embodiment, a computational method is provided comprising: taking an X-ray diffraction pattern of a crystal of the target protein; and computing a three-dimensional electron density map from the X-ray diffraction pattern by molecular replacement, wherein structure coordinates used as a molecular replacement model comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3. This

method may optionally further comprise determining a three-dimensional crystal structure based upon the computed three-dimensional electron density map.

[0027] In one embodiment, a method is provided for evaluating a potential of an entity to associate with a protein comprising: creating a computer model of a protein structure using structure coordinates that comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3. 1; performing a fitting operation between the entity and the computer model; and analyzing results of the fitting operation to quantify an association between the entity and the model.

[0028] In another embodiment, a method is provided for evaluating a potential of an entity to associate with a protein comprising: computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3 Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3; evaluating a potential of an entity to associate with the surface contour by performing a fitting operation between the entity and the surface contour; and analyzing results of the fitting operation to quantify an association between the entity and the computer model.

[0029] In another embodiment, a method is provided for identifying entities that can associate with a protein comprising: generating a three-dimensional structure of a protein using structure coordinates that comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3; employing the three-dimensional structure to design or select an entity that can associate with the protein; and contacting the entity with a protein having at least 65% identity with residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3.

[0030] In another embodiment, a method is provided for identifying entities that can associate with a protein comprising: computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3; employing the computer model to design or select an entity that can associate with the protein; and contacting the entity with a protein having at least 65%, 70, 80, 90, 95% identity with residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3.

[0031] In another embodiment, a method is provided for evaluating the ability of an entity to associate with a protein, the method comprising: constructing a computer model defined by structure coordinates that comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3; selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity. and (iv) modifying a known ligand for DPPIV, or a portion thereof; performing a fitting program operation between computer models of the entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the binding pocket.

[0032] In another embodiment, a method for evaluating the ability of an entity to associate with a protein, the method comprising: computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ

ID NO:3; selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity, and (iv) modifying a known ligand for an DPPIV, or a portion thereof; performing a fitting program operation between computer models of the entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the said binding pocket.

[0035] In another embodiment, a method is provided for identifying an entity that associates with a protein comprising: taking structure coordinates from diffraction data obtained from a crystal of a protein that has at least 65%, 70%, 80%, 90%, 95% or more identity with the residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3; and performing rational drug design using a three dimensional structure that is based on the obtained structure coordinates. The protein crystals may optionally have a crystal lattice having unit cell dimensions, +/- 5%, of a=121.53A b=124.11Å and c=144.42Å, $\alpha=\gamma=90^{o}$, $\beta=114.6^{o}$. The method may optionally further comprise selecting one or more entities based on the rational drug design and contacting the selected entities with the protein. The method may also optionally further comprise measuring an activity of the protein when contacted with the one or more entities. The method also may optionally further comprise comparing activity of the protein in a presence of and in the absence of the one or more entities; and selecting entities where activity of the protein changes depending whether a particular entity is present. The method also may optionally further comprise contacting cells expressing the protein with the one or more entities and detecting a change in a phenotype of the cells when a particular entity is present.

[0038] Figure 3 lists a set of atomic structure coordinates for DPPIV (SEQ ID NO:3) as derived by X-ray crystallography from a crystal that comprises the protein. The following abbreviations are used in Figure 3: "X, Y, Z" crystallographically define the atomic position of the element measured; "B" is a thermal factor that measures movement of the atom around its atomic center; "Occ" is an occupancy factor that refers to the fraction of the molecules in which each atom occupies the position specified by the coordinates (a value of "1" indicates that each atom has the same conformation, i.e., the same position, in all molecules of the crystal). "NAG" stands for N-Acetylglucosamine.

[0043] The present invention relates to a member of the S9 family of human proteases known as dipeptidyl peptidase IV (DPPIV) (SEQ ID NO:1). More specifically, the present invention relates to DPPIV in crystalline form, methods of forming crystals comprising DPPIV, methods of using crystals comprising DPPIV, structure coordinates and a crystal structure of DPPIV, and methods of using the structure coordinates and crystal structure.

[0045] Dipeptidyl Peptidase IV (DPPIV) (SEQ ID NO:1) is a serine protease of Clan SC family S9. DPPIV is a 240kDa homodimeric, multi-functional type-II membrane bound glycoprotein, widely distributed in all mammalian tissues, but highly expressed in kidney, liver and endothelium. DPPIV is also known as DPP4, CD26, adenosine deaminase complexing protein 2 or adenosine deaminase binding protein (ADAbp). DPPIV consists of a short cytoplasmic domain of six amino acids, followed by a hydrophobic transmembrane domain (amino acids 7-28) and an extracellular sequence of 739 amino acids.

[0050] In another embodiment, DPPIV comprises residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3 which comprises the active site domain of wild-type DPPIV that is represented in the set of structural coordinates shown in Figure 3.

[0055] One or more of the sets of amino acids set forth in the tables is preferably conserved in a variant of DPPIV. Hence, DPPIV may optionally comprise a sequence that has at least 65% identity, preferably at least 70%, 80%, 90%, 95% or higher identity with any one of the above sequences (e.g., all of ~~SEQ. ID No. 1~~ SEQ ID NO:3 or residues ~~39-766 of SEQ. ID No. 1~~ 13-740 of SEQ ID NO:3) where at least the residues shown in Tables 1, 2, 3 and/or 4 are conserved with the exception of 0, 1, 2, 3, or 4 residues. It should be recognized that one might optionally vary some of the binding site residues in order to determine the effect such changes have on structure or activity.

Table 1: Amino Acids encompassed by a 4-Angstrom radius around the DPPIV active site (SEQ ID NO:3).

ARG [[111]] 99	TYR [[533]] 521	TYR [[652]] 640
GLU [[191]] 179	SER [[616]] 604	ASN [[696]] 684

GLU [[192]] 180	TYR [[617]] 605	HIS [[726]] 714
SER [[195]] 183	VAL [[642]] 630	ASP [[694]] 682
PHE [[343]] 331	TYR [[648]] 636	

Table 2: Amino Acids encompassed by a 7-Angstrom radius around the DPPIV active site (SEQ ID NO:3).

ARG [[111]] 99	TYR [[533]] 521	TRP [[645]] 633
HIS [[112]] 100	GLY [[535]] 523	TYR [[648]] 636
TRP [[187]] 175	PRO [[536]] 524	ASP [[649]] 637
GLU [[190]] 178	TYR [[571]] 559	TYR [[652]] 640
GLU [[191]] 179	TRP [[615]] 603	THR [[653]] 641
GLU [[192]] 180	SER [[616]] 604	ARG [[655]] 643
VAL [[193]] 181	TYR [[617]] 605	TYR [[656]] 644
PHE [[194]] 182	GLY [[618]] 606	ASN [[696]] 684
SER [[195]] 183	TYR [[620]] 608	VAL [[697]] 685
ARG [[342]] 330	ALA [[640]] 628	HIS [[726]] 714
PHE [[343]] 331	PRO [[641]] 629	ASP [[694]] 682
ARG [[344]] 332	VAL [[642]] 630	

Table 3: Amino Acids encompassed by a 10-Angstrom radius around the DPPIV active site (SEQ ID NO:3).

ARG [[111]] 99	ILE [[391]] 379	SER [[643]] 631
HIS [[112]] 100	VAL [[532]] 520	ARG [[644]] 632
TRP [[187]] 175	TYR [[533]] 521	TRP [[645]] 633
VAL [[188]] 176	ALA [[534]] 522	TYR [[647]] 635
TYR [[189]] 177	GLY [[535]] 523	TYR [[648]] 636
GLU [[190]] 178	PRO [[536]] 524	ASP [[649]] 637
GLU [[191]] 179	CYS [[537]] 525	SER [[650]] 638
GLU [[192]] 180	SER [[538]] 526	VAL [[651]] 639
VAL [[193]] 181	TYR [[571]] 559	TYR [[652]] 640
PHE [[194]] 182	MET [[577]] 565	THR [[653]] 641
SER [[195]] 183	LEU [[584]] 572	GLU [[654]] 642
ALA [[196]] 184	GLU [[588]] 576	ARG [[655]] 643
TYR [[242]] 230	GLY [[614]] 602	TYR [[656]] 644
CYS [[287]] 275	TRP [[615]] 603	MET [[657]] 645
GLN [[306]] 294	SER [[616]] 604	HIS [[690]] 678
TRP [[339]] 327	TYR [[617]] 605	ASP [[694]] 682
VAL [[340]] 328	GLY [[618]] 606	ASP [[695]] 683
GLY [[341]] 329	GLY [[619]] 607	ASN [[696]] 684
ARG [[342]] 330	TYR [[620]] 608	VAL [[697]] 685
PHE [[343]] 331	VAL [[621]] 609	HIS [[698]] 686

ARG [[344]] 332	VAL [[639]] 627	GLN [[701]] 689
PRO [[345]] 333	ALA [[640]] 628	HIS [[726]] 714
SER [[346]] 334	PRO [[641]] 629	GLY [[727]] 715
GLU [[347]] 335	VAL [[642]] 630	

Table 4: Amino Acids encompassed by a 5-Angstrom radius around the AB and CD dimerization interfaces (SEQ ID NO:3).

Chain A	Chain B	Chain C	Chain D
SER A [[225]] 213	PRO B [[220]] 208	PRO C [[220]] 208	LEU D [[221]] 209
TYR A [[227]] 215	ILE B [[222]] 210	LEU C [[221]] 209	ILE D [[222]] 210
SER A [[228]] 216	GLU B [[223]] 211	ILE C [[222]] 210	GLU D [[223]] 211
ASP A [[229]] 217	TYR B [[224]] 212	GLU C [[223]] 211	TYR D [[224]] 212
GLU A [[230]] 218	SER B [[225]] 213	TYR C [[224]] 212	SER D [[225]] 213
LEU A [[232]] 220	TYR B [[227]] 215	TYR C [[227]] 215	SER D [[228]] 216
GLN A [[233]] 221	SER B [[228]] 216	SER C [[228]] 216	ASP D [[229]] 217
TYR A [[234]] 222	ASP B [[229]] 217	ASP C [[229]] 217	GLU D [[230]] 218
PRO A [[235]] 223	GLU B [[230]] 218	GLU C [[230]] 218	SER D [[231]] 219
LYS A [[236]] 224	SER B [[231]] 219	SER C [[231]] 219	LEU D [[232]] 220
THR A [[237]] 225	TYR B [[234]] 222	LEU C [[232]] 220	GLN D [[233]] 221
ARG A [[239]] 227	THR B [[237]] 225	GLN C [[233]] 221	TYR D [[234]] 222
TYR A [[242]] 230	ARG B [[239]] 227	TYR C [[234]] 222	PRO D [[235]] 223
LYS A [[244]] 232	GLN B [[700]] 688	PRO C [[235]] 223	THR D [[237]] 225
ALA A [[245]] 233	ALA B [[703]] 691	THR C [[237]] 225	ARG D [[239]] 227
SER A [[706]] 694	GLN B [[704]] 692	TYR C [[242]] 230	LYS D [[244]] 232
LYS A [[707]] 695	LYS B [[707]] 695	PRO C [[243]] 231	ALA D [[245]] 233
LEU A [[709]] 697	LEU B [[709]] 697	LYS C [[244]] 232	ALA D [[247]] 235
VAL A [[710]] 698	VAL B [[710]] 698	GLU C [[646]] 634	TYR D [[647]] 635
ASP A [[711]] 699	ASP B [[711]] 699	THR C [[673]] 661	MET D [[675]] 663
GLY A [[713]] 701	GLY B [[713]] 701	LEU C [[688]] 676	HIS D [[690]] 678
VAL A [[714]] 702	VAL B [[714]] 702	PHE C [[699]] 687	GLN D [[700]] 688
PHE A [[716]] 704	ASP B [[715]] 703	GLN C [[700]] 688	SER D [[702]] 690
GLN A [[717]] 705	PHE B [[716]] 704	SER C [[702]] 690	GLN D [[704]] 692
MET A [[719]] 707	GLN B [[717]] 705	GLN C [[704]] 692	SER D [[706]] 694
TRP A [[720]] 708	ALA B [[718]] 706	LEU C [[709]] 697	ASP D [[711]] 699
TYR A [[721]] 709		VAL C [[710]] 698	VAL D [[714]] 702
THR A [[722]] 710		ASP C [[711]] 699	ASP D [[715]] 703
ASP A [[723]] 711		GLY C [[713]] 701	PHE D [[716]] 704
		VAL C [[714]] 702	GLN D [[717]] 705
		ASP C [[715]] 703	ALA D [[718]] 706
		PHE C [[716]] 704	MET D [[719]] 707
		GLN C [[717]] 705	TRP D [[720]] 708
		ALA C [[718]] 706	TYR D [[721]] 709
		MET C [[719]] 707	THR D [[722]] 710
		TRP C [[720]] 708	
		TYR C [[721]] 709	